REMARKS

§ 103 REJECTIONS

Claims 1-15 and 17 were rejected under 35 U.S.C. § 103(a) as being unpatentable over Griniasty (U.S. Patent Publication 2003/0088416) in view of Lawrence (U.S. Patent Publication 2003/0049588). Claim 16 was rejected under 35 U.S.C. § 103(a) as being unpatentable over Griniasty in view of Iizuka (U.S. Patent Publication 2001/0009009).

CLAIMS 1-6

Independent claim 1 provides a method of segmenting words into component parts. The method includes determining a mutual information score for a pair of graphoneme units, comprising a first graphoneme unit and a second graphoneme unit. The mutual information score is determined using the probability of the first graphoneme unit appearing immediately after the second graphoneme unit, the unigram probability of the first graphoneme unit and the unigram probability of the second graphoneme unit. Each graphoneme unit comprises at least one letter in the spelling of a word. The mutual information score is used to combine the first and second graphoneme units into a larger graphoneme unit. In a dictionary comprising segmentations of words into sequences of graphoneme units, the first and second graphoneme units are replaced with the larger graphoneme unit in each sequence of graphoneme units in which the first graphoneme unit appears immediately after the second graphoneme unit.

Claim 1 is not shown or suggested in the combination of Griniasty and Lawrence. In particular, neither reference shows or suggests replacing first and second graphoneme units with a larger graphoneme unit in each sequence of graphoneme units in which the first graphoneme unit appears immediately after the second graphoneme unit in a dictionary.

In the Office Action, Lawrence was said to show this limitation in paragraphs 14-24 and 37. Applicants respectfully dispute this assertion. In paragraphs 14-24, Lawrence describes segmenting each word in a dictionary into a sequence of graphonemes. This segmentation is performed by scanning each word in a left to right manner and looking in a table 12 containing clusters of letters and a phoneme used to pronounce those letters. As noted in paragraph 12, table 12 is constructed so that only strings of letters corresponding to a single phoneme are held in the

table. Thus, table 12 contains a list of possible graphonemes with a single phoneme for each graphoneme. In particular, to segment a word, Lawrence begins with the entire word and first looks for the entire word in the table 12. If the entire word is not found as a graphoneme in the table, the last letter of the word is removed and the remaining letters are used to search the table. Letters continue to be removed from the end of the word until a graphoneme is found in table 12. When a graphoneme is found, the remaining letters in the word are then tested in the same manner.

Once the words have been segmented into graphonemes, the individual graphonemes in the dictionary are counted to form a weighting value. The weighting value is then stored for the graphoneme as shown in Fig. 2.

In paragraphs 14-24 of Lawrence, there is mention of replacing first and second graphoneme units with a larger graphoneme unit in each sequence of graphoneme units in the dictionary in which the first graphoneme unit appears immediately after the second graphoneme unit. In fact, there is no mention of replacing any graphoneme units in paragraphs 14-24. Instead, paragraphs 14-24 simply describe how to set an initial segmentation for each word in the dictionary. Lawrence does not say that this initial segmentation is replaced with larger graphoneme units.

In paragraph 37, Lawrence makes reference to silent letters such as "h". In this paragraph, Lawrence indicates that during data collection it is decided to leave "wh" as an orthographic variant for the phoneme "w" rather than have "h" as a silent letter. Applicants note that Lawrence is not stating that a larger graphoneme unit is replacing a first and second graphoneme unit in a sequence of graphoneme units. Instead, Lawrence is indicating that the graphoneme unit consisting of the letters wh and the phoneme w are identified during the selection process. There is no mention of placing any pair of graphoneme units with a larger graphoneme unit.

Since Lawrence does not replace a first graphoneme unit and a second graphoneme unit in each sequence of graphoneme units in a dictionary in which the first graphoneme unit appears immediately after the second graphoneme unit, but instead merely discusses how to segment words in a dictionary into graphoneme units by sequentially identifying the graphoneme units in a left to right manner, the combination of Lawrence and Griniasty does not show or suggest the invention of claim 1 or claims 2-6, which depend therefrom.

CLAIMS 7-15

Independent claim 7 provides a computer-readable storage medium having computer-executable instructions stored thereon that when executed by a processor cause the processor to perform a series of steps. The steps include determining mutual information scores for pairs of graphoneme units found in a set of words. Each graphoneme unit comprises at least one letter. Each mutual information score for a pair of graphoneme units is based on the probability of one graphoneme unit of the pair appearing immediately after the other graphoneme unit of the pair as well as the unigram probabilities of each graphoneme unit in the pair. The graphoneme units of one pair of graphoneme units are combined to form a new graphoneme unit based on the mutual information scores. A segmentation of a word comprising a set of graphoneme units for the word that includes the pair of graphoneme units is updated by replacing the pair of graphoneme units in the segmentation with the new graphoneme unit.

Claim 7 is not shown or suggested in the combination of Griniasty and Lawrence. In particular, neither Griniasty nor Lawrence show or suggest updating a segmentation of a word by replacing a pair of graphoneme units in the segmentation with a new graphoneme unit by combining the pair of graphoneme units.

In the Office Action, paragraphs 14-24 and 37 of Lawrence were said to show updating a segmentation of a word by replacing a pair of graphoneme units in the segmentation with a new graphoneme unit. However, as noted above, the cited paragraphs of Lawrence make no mention of replacing a pair of graphoneme units with a new graphoneme unit. Instead, the cited paragraph merely discusses forming an initial segmentation of words in a dictionary. Lawrence does not indicate that this segmentation involves updating a segmentation of a word comprising a set of graphoneme units for the word by replacing a pair of graphoneme units with a new graphoneme unit. Instead, Lawrence simply describes a single segmentation of each word in the dictionary. There is no mention in any of the cited paragraphs of updating such a segmentation by replacing a pair of graphoneme units in the segmentation with a new graphoneme unit.

Since Lawrence does not update a segmentation of a word by replacing a pair of graphoneme units with a new graphoneme unit, the combination of Lawrence and Griniasty does not show or suggest the invention of claim 7 or claims 8-15, which depend therefrom.

CLAIM 16

Claim 16 provides a method of segmenting a word into syllables. The method includes segmenting a set of words into phonetic syllables using mutual information scores wherein using a mutual information score comprises computing a mutual information score for two phones by dividing the probability of the two phones appearing next to each other in the set of words by the unigram probabilities of each of the two phones appearing in the set of words. The segmented set of words is used to train a syllable n-gram model. The syllable n-gram model is then used to segment a phonetic representation of a word into syllables via forced alignment.

Claim 16 is not shown or suggested in the combination of Griniasty and Iizuka. In particular, neither reference shows or suggests computing a mutual information score by providing a probability of two phones appearing next to each other in a set of words by the unigram probabilities of each of the two phones appearing in the set of words. In the Office Action, it was indicated that Griniasty does not show this mutual information score and that although Iizuka also does not show this exact computation of a mutual information score paragraphs 18, 143, and equations 3, 4, and 6 are obvious variants to provide a method of character string dividing or segmenting. Applicants respectfully dispute this assertion that the mutual information score of claim 16 is obvious from Iizuka.

First, Iizuka does not work with the probabilities of phones. Instead, Iizuka only works with the probabilities of characters. Further, in the cited equations and paragraphs, Iizuka describes a joint probability of a character given a sequence of proceeding characters as being equal to the count of the number of times the sequence of characters including the current character is observed divided by the count of the number of times the preceding characters are observed. Thus, in the division, counts are being divided and not probabilities. Further, the division shown in Iizuka does not involve dividing by the unigram probabilities of each of two phones appearing in the set of words. Instead, at most, Iizuka shows dividing by the count of a single character. Note that the count of two characters appearing next to each other is not a unigram probability since it involves two characters and not a single character.

Since Iizuka forms its probability by counts and not by dividing probabilities and since Iizuka does not divide by unigram probabilities for each of two phones, Iizuka is not an obvious

variant of the mutual information score of claim 16. In particular, the computations performed by Iizuka would provide a different result and have a different meaning than the mutual information score computed in claim 16.

CLAIM 17

Claim 17 provides a method of segmenting a word into morphemes. The method includes segmenting a set of words into morphemes using mutual information scores wherein using mutual information scores comprises computing a mutual information score for two letters based on the probability of the two letters appearing next to each other in the set of words and the unigram probabilities of each of the two letters appearing in the set of words. The segmented set of words is used to train a morpheme n-gram model and the morpheme n-gram model is used to segment a word into morphemes via forced alignment.

Claim 17 is not shown or suggested in the combination of Griniasty and Lawrence. In particular, neither Griniasty nor Lawrence show or suggest computing a score for two letters based on the probability of the two letters appearing next to each other in a set of words and the unigram probabilities of each of the two letters appearing in the set of words. In the Office Action, it was asserted that paragraphs 28 and 37 of Lawrence showed computing a mutual information score based on the probability of two letters appearing next to each other in a set of words and the unigram probabilities of each of the two letters appearing in the set of words. Applicants respectfully dispute this assertion. Under Lawrence, the number of times graphonemes are observed in a dictionary is counted and recorded in a weightings table (see paragraph 17 of Lawrence). Lawrence does not convert these frequencies of occurrence into probabilities. Further, Lawrence does not compute a score based on the probability of two letters appearing next to each other in a set of words and the unigram probabilities of each of the two letters appearing in the set of words. Note, for instance, that in Fig. 2, the letters ou appear twice in the table, once with a weighting of 60,001 and once with a weighting of 960,004. Further, there is a separate weighting for o as 973,566 and u as 20,002. However, Lawrence provides no way of generating a score from the individual weightings for o and u and either of the weightings for ou. Further, the multiple weightings for ou indicates that the weightings provided are not a

probability of two letters next to each other, but instead are a count of the number of times the two letters produce a particular phone.

Since Lawrence does not compute any probabilities of two letters appearing next to each other or unigram probabilities of each of two letters appearing in a set of words and because Lawrence does not provide any way to determine a score based on the probability of two letters appearing next to each other and the probabilities of each of the two letters appearing in the set of words, the combination of Griniasty and Lawrence does not show or suggest the invention of claim 17.

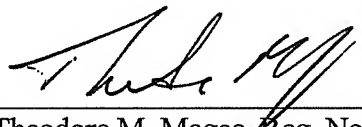
CONCLUSION

In light of the above remarks, claims 1-17 are in form for allowance. Reconsideration and allowance of the claims is respectfully requested.

The Director is authorized to charge any fee deficiency required by this paper or credit any overpayment to Deposit Account No. 23-1123.

Respectfully submitted,

WESTMAN, CHAMPLIN & KELLY, P.A.

By: 

Theodore M. Magee, Reg. No. 39,758
900 Second Avenue South, Suite 1400
Minneapolis, Minnesota 55402-3319
Phone: (612) 334-3222 Fax: (612) 334-3312

TMM:kjl